

TRACKING TRACE'S TROUBLES

Anne Cutler

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands.

Dennis Norris

MRC Cognition and Brain Sciences Unit, Cambridge, U.K.

and James McQueen

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands.

ABSTRACT

Simulations explored the inability of the TRACE model of spoken-word recognition to model the effects on human listening of acoustic-phonetic mismatches in word forms. The source of TRACE's failure lay not in its interactive connectivity, and not in the presence of inter-word competition, and not in the use of phonemic representations, but in the need for continuously optimised interpretation of the input. When an analogue of TRACE was allowed to cycle to asymptote on every slice of input, an acceptable simulation of the subcategorical mismatch data was achieved. Even then, however, the simulation was not as close as that produced by the Merge model.

1. INTRODUCTION

A major distinction among models of the recognition of spoken words is whether or not a model allows feedback from logically later to logically earlier levels of processing. One of the leading current models is TRACE [3]. TRACE is an interactive model which allows feedback from words to sublexical representations. Other models (e.g. Shortlist [5]) do not allow such feedback. Figure 1 sketches the differences between these models in the connection architecture; the arrows in the sketch represent flow of information, and the lines ending with filled circles represent the existence of inhibitory connections between nodes within a given level of the model. The important difference between the two models for present purposes is the downward arrow allowing flow of information from the word level to the phoneme level in TRACE; no such connection exists in Shortlist.

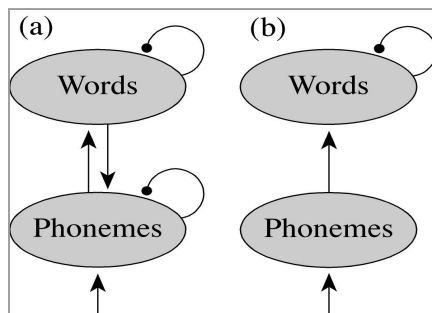


Figure 1. Connections between sublexical and lexical levels in TRACE (a) and Shortlist (b).

Recent research in spoken-word recognition [2, 4] has posed a serious challenge to TRACE. In this research, listeners made phonemic decisions or lexical decisions on words and nonwords, some cross-spliced so that they contained acoustic-phonetic mismatches. For instance, the word *job* could have its initial portion *jo-* taken from the word *jog* or the nonword *jo-*; the nonword *smob* could have its initial portion *smo-* taken from the word *smog* or the nonword *smo-*. In such cross-spliced items, a "subcategorical mismatch" [9] arises when the initial portion originally had a different following consonant; formant transitions in the vowel signal the original following consonant, and these are contradicted by the consonant which actually follows.

Marslen-Wilson and Warren [2] found that both phonemic decisions and lexical decisions were slower for mismatching items than for matching items (a matching item has, for example, *jo-* from *job* added to *-b* from another *job*). This was in agreement with earlier findings [9]. However, Marslen-Wilson and Warren also observed a difference between the word and nonword items. Whereas responses (in both phonemic decision and lexical decision) were equivalently slowed for both types of mismatching cross-spliced words in comparison to the matching words, an asymmetry arose with the nonwords: there was a significantly greater response disadvantage for nonwords cross-spliced with words than for nonwords cross-spliced with other nonwords.

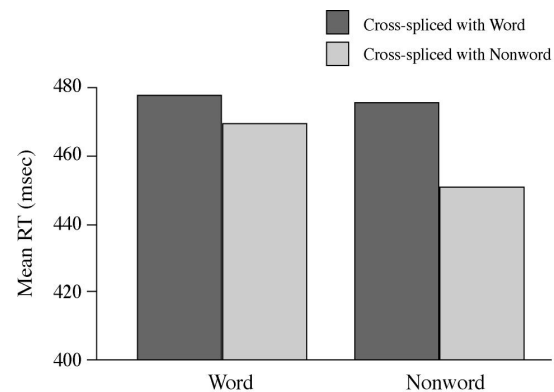


Figure 2. Mean "YES" reaction time (RT) to words and "NO" RT to nonwords in lexical decision; data from [4].

McQueen, Norris and Cutler [4] replicated and extended the Marslen-Wilson and Warren study, adding an additional task (phoneme detection), and demonstrating that the asymmetric disadvantage in the

mismatching nonwords could be made to come and go as a function of task demands. Figure 2 shows the relevant portion of the lexical decision data from McQueen et al.'s study. It can be seen that "YES" responses to words cross-spliced with words and to words cross-spliced with nonwords were not significantly different, but "NO" responses were slower to nonwords cross-spliced with words than to nonwords cross-spliced with nonwords. The same asymmetry appeared in the phoneme decision data of one of McQueen et al.'s experiments.

2. SIMULATIONS WITH UNADAPTED MODELS

The standard version of TRACE cannot simulate these findings. Marslen-Wilson and Warren [2] conducted simulations with standard TRACE, and showed that the model was unable to account for their data in several respects. First, the response probabilities at the lexical level were computed in an attempt to simulate lexical decision. Here TRACE showed a large effect of the source of a mismatching cross-splice for words as well as for nonwords, which was not the case in the human data. For the nonword stimuli the nonwords cross-spliced with words items furthermore produced response probabilities which were as large as the probabilities for any of the word stimuli. These nonwords should therefore have been systematically misclassified as words; human listeners, however, did not in fact make this error. Second, the phonetic categorization results were simulated by using response probabilities calculated at the phoneme level. As in the lexical decision simulations, TRACE incorrectly showed a large effect of the source of a mismatching cross-splice for words as well as for nonwords. This predicted that listeners should have had more difficulty with words cross-spliced with words than with words cross-spliced with nonwords, but again, this was not the case.

Marslen-Wilson and Warren suggested that the failure of TRACE to simulate their data might be attributable to the model's use of lateral inhibition and top-down feedback, and to the fact that it does not use mismatch information. Furthermore, they claimed that TRACE could be viewed as "the only viable candidate of the classical representational type" (p. 673), by which term they denoted a model using phonemes rather than features as sublexical representations, and they argued that the simulation failure thus constituted a general argument against the whole class of models in which phonemes play a role as prelexical representations.

However, these suggestions raised by Marslen-Wilson and Warren cannot be the reasons why TRACE failed to simulate the data. Another model which has both phonemic representations and inter-word competition can happily simulate the same data. This is the Merge model, a model of phonemic decision-making which is integrated with Shortlist [4, 6]. Merge is an autonomous model, i.e. it does not allow the flow of information from the lexical to the sublexical level as in TRACE (Figure 1). In Merge, phonemic decisions are made by a dedicated decision-making process which

accepts information from both phonetic processing and lexical activation. The model is implemented as a simple competitive network model, and several simulations with the model are reported by Norris et al. [6].

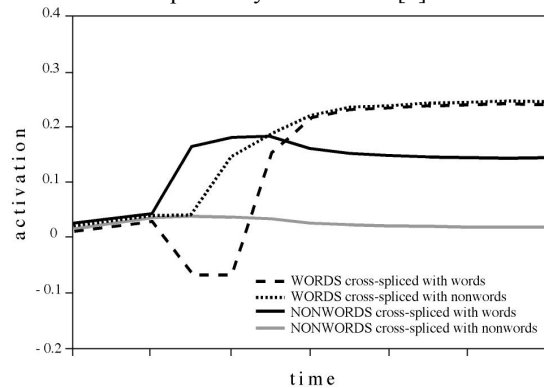


Figure 3. Activation levels in Merge simulations of the lexical decision data in Figure 2. Activation is shown for the relevant lexical nodes for each condition.

Figure 3 shows a Merge simulation of the lexical decision data depicted in Figure 2. Activation levels across time can be compared for the relevant lexical nodes associated with the four different types of mismatching cross-spliced item. That is, for word items, the activation level is shown for the word for which a "YES" response is made. It can be seen that the two different types of mismatching word (dotted and dashed lines) in fact reach the same level at asymptote, which is consistent with the result of the experiment (Figure 2), in which there was no significant difference in these "YES" responses. For the nonword items, activation is plotted for the lexical item which provided the word-onset which was cross-spliced - i.e., *smog* in the example given above. In the case of nonwords cross-spliced with other nonwords (e.g. *smob* in which *smo-* came from *smod*; grey line), there is no competing lexical activation, but in the case of nonwords cross-spliced with words (e.g. *smob* in which *smo-* came from *smog*; solid line), there is significant activation, again consistent with the result from the experiment (Figure 2), in which "NO" responses to nonwords cross-spliced with words were delayed relative to the other mismatched nonwords. The activation does not rise as high as that for the real words, however, consistent with the fact that the listeners did not erroneously classify these nonwords as real words. For greater detail of this simulation see Norris et al. [6].

3. MODEL ADAPTATIONS

So why can Merge simulate the data while TRACE cannot? Although Merge and Shortlist are like TRACE in having sublexical phonemic representations and inter-word competition, they also differ from TRACE in a number of ways. First, TRACE incorporates feedback between processing levels, while Merge and Shortlist do not. Second, in Merge/Shortlist, the inter-word competition produces a continuously optimal lexical parse of the input, but this is not the case in TRACE.

We explored the reasons why TRACE could not simulate the crucial data. Using a small-scale version of TRACE, we systematically altered features of the model and compared the performance of each version with the successful Merge simulation. This small-scale TRACE analogue was, essentially, a modified version of the Merge model. We eliminated Merge's decision nodes, and we added feedback from the word to the phoneme layer and within-level inhibition at the phoneme layer. By these means we transformed a Merge network into an interactive model with the same connectivity pattern as TRACE. This enabled us to investigate whether some simple manipulation might enable this TRACE analogue to simulate the subcategorical mismatch data after all.

The simplest kind of interactive model we could try was one with dynamics like the original TRACE model: one network cycle per time slice and no phoneme-word inhibition and no resetting of activations. For these simulations, as described, there was word-to-phoneme feedback and within-level inhibition at the phoneme level. In TRACE the between-phoneme inhibition is required in order to reach an unambiguous decision as to which phoneme is in the input. Because there was between-phoneme inhibition in the TRACE analogue, the phoneme level cycled in synchrony with the word level.

Figure 4 shows the result of a simulation which yielded the same poor fit as that obtained by Marslen-Wilson and Warren with the full TRACE model. The model incorrectly predicted a difference between the two types of mismatching cross-spliced word: it can be seen that the lines representing the two word types do not reach the same level of activation. The model also exaggerated the difference between the two types of mismatching cross-spliced nonword, to the extent that nonwords cross-spliced with words produced as much lexical activation as words cross-spliced with words. This would predict the same pattern of response for both item types, i.e. a very high error rate for the nonwords cross-spliced with words (i.e. many incorrect "YES" responses to *smob* when the *smo-* came from *smog*). There was no such effect in the experimental data, as pointed out above; but Marslen-Wilson and Warren's simulations, and the present ones with the TRACE analogue, both produced this unwanted effect.

We made extensive attempts to improve this model's fit to the data by setting parameters by hand, but these attempts were completely unsuccessful. In fact, it was hard to be sure exactly why it was proving so difficult to discover a suitable set of parameters. Parameter setting in an interactive model is, by the nature of the model, very difficult to do, since adjustments to the phoneme parameters alter lexical behavior and vice versa. We therefore decided to investigate use of an optimization procedure to set the parameters of the interactive model automatically in order to reproduce the same activation pattern as the autonomous model, Merge (see Figure 3). For this we used Powell's conjugate gradient descent method [7] to fit the parameters of the interactive model to target parameters of Merge.

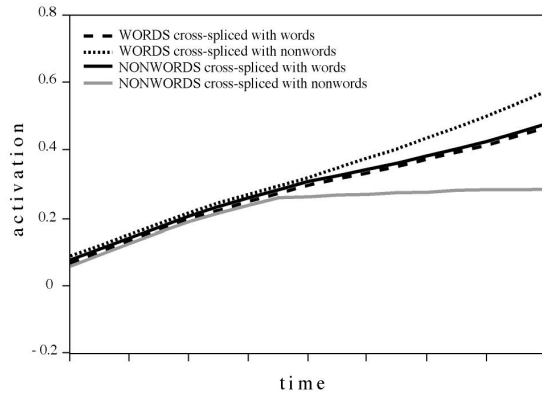


Figure 4. Activation levels in TRACE simulations of the lexical decision data in Figure 2. Activation is shown for the relevant lexical nodes for each condition.

In these optimizations, the phoneme activations of the TRACE analogue were set to reproduce the decision unit activations of Merge. Correlations of lexical and phonemic activations were computed independently and we attempted to maximize the sum of those two correlations. Note that even this was unsuccessful when we used the best Merge simulation as target. Better results were obtained using another Merge parameter set which was not optimal but which still allowed Merge to give a plausible account of the data. The TRACE simulation was required to produce the same pattern of activation as Merge did regardless of any differences in absolute activation levels. These optimization procedures did improve the model's fit to the data, but it was impossible to find a single set of parameters to fit both the lexical decision and the phoneme decision data.

We next constructed a range of networks, each successively more similar to Merge, but all with TRACE's architecture (single-outlet, and feedback from lexical to phonetic processing). We increased the number of cycles per slice and added resetting activation after each time slice. In brief, we found that the more closely the model resembled Merge, the better it could simulate the data. Eventually we found a version of the model that gave an acceptable simulation. However, the model was very unstable and even the very best version never produced as close a simulation of the data as Merge did.

The simulation by our best TRACE analogue is shown in Figure 5. In this simulation, the model used 15 cycles per slice, a momentum term at both the word and phoneme levels, reset and no bottom-up inhibition. For the momentum term, some proportion of the final activation level at the end of the previous time slice was added to the node's input at each cycle. The parameters used in this simulation are given in an appendix to Norris et al. [6]. This model clearly provides a reasonable fit to the lexical decision data, quite comparable with the Merge simulation results plotted in Figure 3 above. However, it should be noted that again this TRACE analogue could not give as good a representation of the phonemic decision data as the Merge simulation did, since the TRACE analogue showed much larger word-nonword differences than appeared the human data.

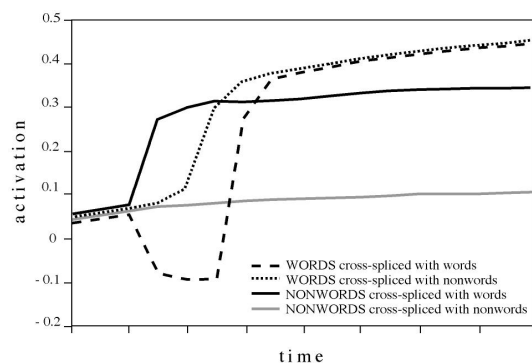


Figure 5. Activation levels in simulations of the lexical decision data in Figure 2 with the best TRACE analogue. Activation is shown for the relevant lexical nodes for each condition. The results are very similar to Figure 3.

4. CONCLUSION

The simulations presented here have confirmed that models incorporating phonemic representations and lexical competition can simulate the detailed pattern of results seen in subcategorical mismatch studies. Marslen-Wilson and Warren's explanations of their failure to simulate the data with TRACE are thus invalid. Further, it is not material whether the models producing such successful simulations are autonomous (like Merge) or interactive (like TRACE).

The best-performing small-scale TRACE analogue differed from Merge principally by having feedback. Thus since this TRACE analogue, like Merge, can simulate the data, the presence of feedback could not be the reason for the failure of the simulation by the standard version of TRACE. Instead, the crucial feature which the TRACE adaptation required was the addition of the Shortlist continuous optimisation procedure, involving 15 cycles of interactive activation per time slice. That is, the primary reason why TRACE is unable to account for the subcategorical mismatch findings is that it does not allow lexical level processes to cycle to asymptote on a small enough time scale. As discussed above, the model thus incorrectly predicts competition effects in the words cross-spliced with words, and the same problem probably also causes TRACE to overestimate inhibition in the nonwords cross-spliced with words.

We were quite unable to find a set of parameters for a single-cycle version of the TRACE analogue which produced a plausible simulation of the data. A single cycle is insufficient to allow the winning lexical candidate to completely suppress its competitors, and so eliminate the competitor effect for words. Note that for both the TRACE-like and Merge 15-cycle simulations, the levels of activation for words at asymptote are almost completely independent of the presence of a competitor. (Any residual competitor effect at asymptote is eliminated completely by using 30 cycles per slice.)

Although we have shown that TRACE can be adapted to simulate the subcategorical mismatch data, it should be noted that TRACE remains quite inconsistent

with compensation for coarticulation data reported by Pitt and McQueen [7]. Furthermore, it still suffers from the limitation first raised by Cutler, Mehler, Norris and Segui [1] in the context of the facilitatory effects of lexical knowledge on phoneme identification; these come and go with modulations of the experimental situation. McQueen et al. [4] showed that the subcategorical mismatch effect comes and goes according to the nature of the task and the stimulus materials. Although one could stipulate that the top-down connections in TRACE be modulated according to the experimental situation, it is unclear why an interactive model should sometimes choose to forgo the supposed benefits of top-down feedback; thus these demonstrations of variable effects continue to cause problems for TRACE or other feedback models.

The Merge model of Norris et al. [6], on the other hand, is a dual-outlet model where the decision mechanism can selectively emphasize either phonemic or lexical knowledge without in any way altering the bottom-up nature of the word recognition process itself. This architecture naturally copes with variable effects as a function of task and materials. Moreover, as we have shown (and for more detail see Norris et al. [6]), the Merge model simulates the present set of results better than any achievable version of TRACE can manage.

5. REFERENCES

- [1] Cutler, A., Mehler, J., Norris, D.G. & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.
- [2] Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, **101**, 653-675.
- [3] McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- [4] McQueen, J.M., Norris, D.G. & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 1363-1389.
- [5] Norris, D.G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.
- [6] Norris, D.G., McQueen, J.M. & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, **23**.
- [7] Pitt, M.A. & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, **39**, 347-370.
- [8] Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
- [9] Whalen, D.H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.